

# What are the measures that can be used to assess performance during in situ Paediatric Emergency Medicine Simulation?

Jennifer Amanda Mann,<sup>1,2</sup> Damian Roland<sup>1,2</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjstel-2016-000140>).

<sup>1</sup>Paediatric Emergency Medicine Leicester Academic (PEMLA) Group, Paediatric Emergency Department, Leicester Royal Infirmary, Leicester, UK  
<sup>2</sup>SAPPHIRE group, Health Sciences, Leicester University, Leicester, UK

## Correspondence to

Dr Jennifer Amanda Mann, Paediatric Emergency Medicine Leicester Academic (PEMLA) Group, Paediatric Emergency Department, Leicester Royal Infirmary, Infirmary Square, Leicester LE1 5WW UK; [jenniferhowes@doctors.org.uk](mailto:jenniferhowes@doctors.org.uk)

Received 15 July 2016  
Revised 7 December 2016  
Accepted 17 February 2017  
Published Online First  
20 March 2017

## ABSTRACT

**Background** Paediatric in situ simulation within emergency departments is growing in popularity as an approach for improving multidisciplinary team working, enabling clinical skills development and exploring the importance of human factors in the clinical setting. However, measuring the success of such programmes is often through participant feedback of satisfaction and not measures of performance, which makes it difficult to assess whether such programmes lead to improvements in clinical behaviour.

**Objective** To identify the measures that can be used to assess performance during in situ paediatric emergency medicine simulations.

**Study selection** A literature search of EMBASE, ERIC and MEDLINE was performed using the key terms (Paediatrics and Emergency and Simulation.) MeSH and subheadings were used to ensure all possible variations of the key terms were included within the search.

**Findings** The search revealed 607 articles, with 16 articles meeting inclusion criteria. Three themes of evaluation strategy were identified—the use of feedback forms (56% n=9/16), performance evaluation methods (63% n=10/16) or other strategies (25% n=4/16), which included provider comfort scores, latent safety threat identification and episodes of suboptimal care and their causation.

**Conclusions** The most frequently used method of assessment in paediatric emergency department simulation are performance evaluation methods. None of the studies in this area have looked at patient level outcomes and this is therefore an area which should be explored in the future.

## BACKGROUND

The use of simulation is growing exponentially within paediatrics; ranging from structured courses to departmental training days and the running of simulation scenarios within units. In paediatrics, simulation has been used to assess a variety of clinical scenarios including resuscitation, trauma, airway skills, procedural techniques and crisis management.<sup>1</sup> Simulation allows the development of skills in a non-clinical setting, rather than with real patients and therefore, clinical errors can be used as a learning tool, rather than leading to detrimental clinical outcomes.<sup>2</sup>

Paediatric emergency departments in the UK are now running their own in situ simulations, coordinated by paediatric emergency medicine consultants and other team members, which may include education fellows and clinical skills teams. The aim of these scenarios is to create a realistic case from

which the group can develop team building as well as educational learning outcomes. The Royal College of Paediatric and Child Health created a simulation research subgroup to quantify the extent and content of child health relevant simulation research that has been undertaken in the UK in the previous decade.<sup>3</sup> Conclusions drawn by the group were that while there are a large variety of educational outcomes measured during simulations, often only Kirkpatrick 1 and 2 outcomes are being assessed.<sup>4</sup> This participant feedback commonly takes the form of informal measures of user satisfaction, which can make it difficult to assess whether simulation programmes are successful in bringing about improved performance.

## OBJECTIVE

To identify the measures that can be used to assess performance during in situ paediatric emergency medicine simulations.

## STUDY SELECTION

Working with a senior clinical librarian, a literature search of EMBASE, ERIC and MEDLINE was performed using the key terms (Paediatrics and Emergency and Simulation.) MeSH and subheadings were used to ensure all possible variations of the key terms were included within the search. The search history is included as an online supplementary material file. One author searched through the abstract results to select suitable articles. Inclusion criteria were articles describing in situ paediatric emergency medicine simulations either performed in paediatric emergency departments or run by paediatric emergency department consultants. Commentaries, conference abstracts and external simulation programmes and courses were excluded.

The search revealed 607 articles of which 18 articles met the criteria for inclusion. On full paper review of these articles, a further 2 were excluded leaving the remaining 16 articles. The reasons for the two exclusions were because the full text revealed that the study was a continuation of a previous included study<sup>5</sup> and in the other, candidates were medical students.<sup>6</sup>

## FINDINGS

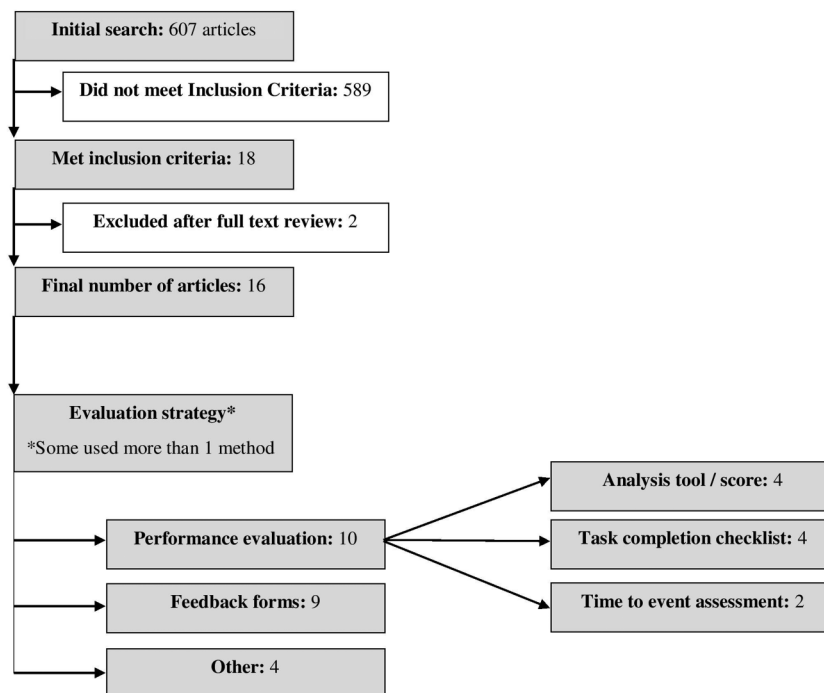
A flow diagram of the literature search outcomes can be found in [figure 1](#). Ten of the articles originated from the US, two from Australia and one each from Canada, Germany, Switzerland and Taiwan. Three themes of evaluation strategy were identified in these articles—the use of performance



CrossMark

To cite: Mann JA, Roland D. *BMJ Stel* 2017;**3**:83–87.

**Figure 1** Flow diagram of the literature search outcomes.



evaluation methods (63%  $n=10/16$ ), feedback forms (56%  $n=9/16$ ) or other strategies (25%  $n=4/16$ ), which included provider comfort scores, latent safety threat identification and episodes of suboptimal care and their causation. Five of the articles used two evaluation strategies and in one article all three strategies were used. We used the Mixed Methods Appraisal Tool<sup>7</sup> to ensure all the studies were of high quality, with all the included studies scoring 75–100%. A table showing the scores for each study as well as the evaluation strategy, age range of the simulated patient and limitations as stated by the author can be found in the online supplementary appendix.

### Feedback forms

Articles using feedback forms universally followed a Likert scale format; however, the focus of the data collection varied from self-perceived reports of performance and preparedness,<sup>8</sup> to clinical impact,<sup>9–10</sup> user satisfaction,<sup>11–12</sup> confidence<sup>13</sup> and knowledge acquisition.<sup>2–14</sup> Feedback form response rates varied from 54 to 100% and were typically completed immediately after the simulation session.

The exception was Happel *et al*,<sup>8</sup> who audited residents after live critical events, regarding perceptions of previous simulations. Findings here were derived from 47 surveys completed by 20 paediatric residents in relation to 27 critical events. Residents reported that their experiences in preceding similar simulations positively affected their performances during actual clinical events. However, there was no statistically significant change in confidence levels between those who reported having had a preceding similar simulation and those who had not. The author recognised the limitation in their retrospective collection of perceptions of previous simulations, stating recall bias may have positively or negatively affected the perceived simulation training effectiveness and response rate to questionnaire may have been affected by more engaged residents responding to questionnaires selectively.

### Performance evaluation methods

Performance evaluation methods were the most commonly used strategy in the identified articles. Of these, 40% ( $n=4/10$ ) used

analysis tools/scores,<sup>10–11–13–15</sup> 40% ( $n=4/10$ ) task completion checklists<sup>16–19</sup> and in two articles time to event assessment.<sup>2–20</sup>

### Analysis tools/scores

Auerbach *et al*<sup>11</sup> used a Trauma Simulation Evaluation tool to analyse the performance of 398 members of their multidisciplinary trauma team in 22 simulations. Over the course of the two years, the authors found statistically significant upward trends in overall performance, intubation and teamwork components. However, the origins of weighting for the individual components of the scoring system were not stated within the paper.

Tsai *et al*<sup>13</sup> designed a scoring system derived from task-specific technical skill, medication and behavioural scores. Couto *et al*<sup>15</sup> used the TEAM assessment score to perform a three-way comparison of multidisciplinary teamwork performance in actual paediatric emergencies, in situ simulations and in centre simulations. The study involved reviews of 132 video recordings (44 in each category) by two expert reviewers. Steps were taken to ensure drift in scoring was limited between reviewers by a process of dual review and discussion of discrepancies after every 10 recordings. The study found similar scores in all three environments.

Zimmerman *et al*<sup>10</sup> describe the impact of implementing inter professional simulation training in their department using a combination of participant feedback and identification of latent safety threats. They also used TeamMonitor, a teamwork self-assessment-based tool to record participant thoughts following live critical events; however, this was not used to assess the simulation component directly.

Review of these articles shows these tools vary greatly in their assessment criteria—Couto *et al*<sup>15</sup> and Zimmerman *et al*<sup>10</sup> focused on teamwork performance, while in addition, the tools used by Auerbach *et al*<sup>11</sup> and Tsai *et al*<sup>13</sup> scored participants for more case specific aspects of medical assessment and management. Global rating scores were also used by Auerbach *et al*<sup>11</sup> and Couto *et al*<sup>15</sup> with measures stated to limit inter-rater variability. These scores were weighted heavily, accounting for 21–27%<sup>11</sup> (depending on whether intubation performed as

an additional skill component) and 48%<sup>15</sup> (dependent on evaluation technique used) of the total points available, respectively.

### Task completion checklists

The development of checklists identified in this search has been carefully considered—with all authors ensuring due attention to the validity of components.

Greenberg *et al*<sup>16</sup> used a checklist for topics to cover when breaking bad news of infant and child death to parents. The elements of the checklist were created following a nationwide survey of emergency medicine physicians and showed improvements in scores for areas of communication and required follow-up.

Hunt *et al*<sup>17</sup> performed a large scale study across 25 emergency departments in North Carolina, running unannounced in situ trauma simulations and measuring adherence to a 44-item checklist with items derived from resuscitation courses that are considered the standard of care for the treatment of critically ill children and trauma patients in the USA.

Schmutz *et al*<sup>18</sup> designed three performance evaluation checklists for simulated cases of infants with cardiopulmonary arrest, dyspnoea with low oxygen saturations after intubation and respiratory syncytial virus bronchiolitis. Elements of the checklist were derived from clinical guidelines and expert opinions through a Delphi process. They assessed construct validity by using three external constructs—global performance rating, team experience level and scenario specific time markers. These performance evaluation checklists were considered valid if they showed a significant relation to the performance score; however, this relationship was not observed in the bronchiolitis case.

Donoghue *et al*<sup>19</sup> addressed concerns of the insensitivity of checklist approaches to the timeliness and order of observed action. They assessed 20 participants in their management of simulated respiratory and cardiac emergency scenarios, using checklists based on Paediatric Advanced Life Support algorithms. Inter-rater reliability for the four raters was calculated to be acceptable. As stated by the author, the goal of the rating systems was to assess 'whether tasks were performed at all, whether they were performed properly, whether they were performed in proper sequence and whether they were performed in a timely manner'.<sup>19</sup>

### Time to event assessment

In 2013, O'Leary *et al*<sup>2</sup> performed a two-arm cohort study with 56 junior medical officers, randomising them to one of two simulated anaphylaxis scenarios; a patient presenting in anaphylaxis with or without associated hypotension. They then used time to event assessment to measure the time epinephrine was given and also the route and dose as secondary outcomes. Although they found statistically significant differences between the two groups, they highlighted that doctors may perform differently in real patient situations which is a limitation of this method of assessment. Tofil *et al*<sup>20</sup> conducted a randomised cohort study to assess whether repeated exposure to one simulation scenario (pulseless electrical activity cardiac arrest) would translate into improved performance and decision-making in varied scenarios. Results showed repetition of exposure did improve some measure of performance in the repeated scenarios. However, conversely a noted limitation was that this influence from previous simulation exposure may not lead to the appropriate prioritisation for the current scenario.

### Other methods

Patterson *et al*<sup>9</sup> and Zimmermann *et al*<sup>10</sup> used latent safety threat identification during debriefs. This brought about change in the form of new clinical developments within their departments; renewing guidelines, setting up workshops and purchasing new equipment. O'Leary *et al*<sup>21</sup> used a similar approach, using simulation to capture episodes of suboptimal care and their causation. Katznelson *et al*<sup>22</sup> used provider comfort scores, recording participants self-rated confidence score as a percentage for procedural skills (including intravenous line placement and blood taking) and for two cognitive skills—patient assessment and recognising abnormal vital signs. Of note, no statistically significant change was found in the reported confidence of medical doctors performing procedural skills prestimulation and postsimulation.

## CONCLUSIONS

### Feedback form

The main caveat with a feedback form is they are not an objective measure of performance within the simulation and second, self-reports of improved confidence do not necessarily translate into improved competence in the clinical context.<sup>23</sup> Self-assessment is a complex, potentially learned skill, requiring individuals to have insights into their own limitations and competencies.<sup>24</sup> This is also supported by previous studies which have shown no relationship between self-ratings of confidence and actual competence.<sup>25</sup>

### Performance evaluation

#### Analysis tools/scores

One of the benefits of an analysis tool is a definitive score you can assess for change over time following exposure to simulation<sup>11</sup> or introduction of other education strategies. Documenting scores for each subdivision also highlights areas to discuss during debriefs. However, completing the tool is often time consuming and it is difficult to set a threshold for what score denotes a level for acceptable clinical practice. Paediatric emergencies are very varied in their presentation and complexities. Running alongside clinical practice, the skill mix and training level of staff participating in an in situ simulation, is dependent on the activity within the department. If an analysis tool which generates an overall score is used, these confounders could generate a low score in a complex scenario with more junior team members and it may disempower them or discourage future participation. Ensuring consistency of global rating scores would also be difficult to achieve in departments which use different facilitators depending on staff availability.

### Task completion checklists

Benefits of a checklist approach are that there is a clear record of which aspects of assessment or management have been missed and in a team or individual who is performing well with minor omissions; this can be very useful in fine tuning their future approach. However, in a complex scenario, the checklist can be very extensive and the importance of a particularly crucial management point overlooked. Regehr *et al*<sup>26</sup> compared the use of checklists and global rating scores in the medical Observed Structured Clinical Examination assessment of 33 medical clerks and 15 residents using simulated patients. Here it was suggested that stations relying only on checklists, rewarded candidate thoroughness rather than competence and may not allow for alternative approaches to a clinical problem which

come with experience. The results supported their initial hypothesis, with global rating scores showing residents scoring statistically significantly higher than clerks, while checklist scores were unable to differentiate between the two.

This work is further supported by Ilgen *et al.*,<sup>27</sup> who performed a systematic review of validity evidence for checklists versus global rating scales in a simulation-based assessment of health professionals. Conclusions were that while checklist inter-rater reliability and trainee discrimination were more favourable than suggested, each task requires a separate checklist. Compared with the checklist, the Global Rating Scale (GRS) has higher average interitem and interstation reliability, can be used across multiple tasks and may better capture nuanced elements of expertise.

### Time to event assessment

Selecting a key part of the management plan and timing this to produce an objective result seems a much more simple method for assessing performance. There are time recommendations for certain paediatric medical treatments, for example, intravenous antibiotics within the first hour of suspected sepsis,<sup>28</sup> which lend themselves well to this method of assessment. However, in the majority of cases, it is the chronology of management and the thought process behind this which is more important. For example, in those performing a structured A to E assessment, intravenous fluids will be given after ensuring that the airway is patent. This means that those with a more haphazard, unstructured approach may score better on 'time to intravenous fluid administration' even though the patient remains at risk with a threatened airway.

### Other methods

Evaluation by latent safety threat identification or episodes of suboptimal care and causation factors provides a good clinical quality improvement correlation, so long as there is a system in place to bring about change and repeat scenarios to see if recognition and avoidance of such issues improves after these changes.

Departments using high-fidelity manikins do so to make their scenarios as realistic as possible and allow for both training of practical skills as well as learning gained from the simulation itself. Intubation, venepuncture and intraosseous needle insertion can be practiced as part of a simulation although the context of this training is different from when occurring in a one-to-one task only session. This may explain why Donoghue *et al.*<sup>19</sup> found no statistically significant difference in self-rated confidence scores for these procedures.

### Summary

There are 16 articles assessing performance in paediatric emergency department in situ simulation, with the most frequently used method of assessment being performance evaluation methods. For a tool to be useful in this environment, it must be easy to use by a variety of facilitators and versatile, reflecting the wide diversity of paediatric emergencies. If timed elements are used, they have the most educational value when viewed in conjunction with the sequential assessment and management of the patient. Exploring the verbal and non-verbal communication within the team must also be considered although this is difficult to quantify. None of the studies in this area have looked at patient level outcomes and this is therefore an area which should be explored in the future.

**Twitter** Follow Damian Roland @damian\_roland

**Acknowledgements** The authors would like to acknowledge Sarah Sutton, Senior Clinical Librarian, Leicester Royal Infirmary for her assistance with this literature search.

**Contributors** JAM and DR planned the review. The literature search was conducted by JM. JAM selected suitable articles which met the inclusion criteria. JAM and DR critically appraised the articles included in the review and co-wrote the review.

**Competing interests** None declared.

**Provenance and peer review** Not commissioned; externally peer reviewed.

### REFERENCES

- Weinberg ER, Auerbach MA, Shah NB. The use of simulation for pediatric training and assessment. *Curr Opin Pediatr* 2009;21:282–7.
- O'Leary FM, Hokin B, Enright K, *et al.* Treatment of a simulated child with anaphylaxis: an in situ two-arm study. *J Paediatr Child Health* 2013;49:541–7.
- Roland D, Wilson H, Holme N, *et al.* Developing a coordinated research strategy for child health-related simulation in the UK: phase 1. *BMJ Simul Technol Enhanced Learn* 2015;1:40.
- Kirkpatrick D, Kirkpatrick J. *Evaluating training programs: the four levels*. 3rd edn. San Francisco: Berrett-Koehler Publishers, Inc., 2006.
- Hunt EA, Heine M, Hohenhaus SM, *et al.* Simulated pediatric trauma team management: assessment of an educational intervention. *Pediatr Emerg Care* 2007;23:796.
- Fielder EK, Lemke DS, Doughty CB, *et al.* Development and assessment of a pediatric emergency medicine simulation and skills rotation: meeting the demands of a large pediatric clerkship. *Med Educ Online* 2015;20:29618.
- Pluye P, Robert E, Cargo M, *et al.* Proposal: a mixed methods appraisal tool for systematic mixed studies reviews. 2011. Retrieved on [21 June 16] from: <http://mixedmethodsappraisaltoolpublic.pbworks.com>. Archived by WebCite® at: <http://www.webcitation.org/5tTRtc9yl>
- Happel CS, Lease MA, Nishisaki A, *et al.* Evaluating simulation education via electronic surveys immediately following live critical events: a pilot study. *Hosp Pediatr* 2015;5:96.
- Patterson MD, Geis GL, Falcone RA, *et al.* In situ simulation: detection of safety threats and teamwork training in a high risk emergency department. *BMJ Qual Saf* 2013;22:468.
- Zimmermann K, Holzinger IB, Ganassi L, *et al.* Inter-professional in-situ simulated team and resuscitation training for patient safety: description and impact of a programmatic approach. *BMC Med Educ* 2015;15:189.
- Auerbach M, Roney L, Aysseh A, *et al.* In situ pediatric trauma simulation: assessing the impact and feasibility of an interdisciplinary pediatric in situ trauma care quality improvement simulation program. *Pediatr Emerg Care* 2014;30:884–91.
- Cheng A, Goldman RD, Aish MA, *et al.* A simulation-based acute care curriculum for pediatric emergency medicine fellowship training programs. *Pediatr Emerg Care* 2010;26:475.
- Tsai TC, Harasym PH, Nijssen-Jordan C, *et al.* Learning gains derived from a high-fidelity mannequin-based simulation in the pediatric emergency department. *J Formos Med Assoc* 2006;105:94.
- Auerbach M, Kessler D, Foltin JC. Repetitive pediatric simulation resuscitation training. *Pediatr Emerg Care* 2011;27:29–31.
- Couto TB, Kerrey BT, Taylor RG, *et al.* Teamwork skills in actual, in situ, and in-center pediatric emergencies: performance levels across settings and perceptions of comparative educational impact. *Simul Healthc* 2015;10:76–84.
- Greenberg LW, Ochenschlager D, O'Donnell R, *et al.* Communicating bad news: a pediatric department's evaluation of a simulated intervention. *Pediatrics* 1999;103:1210.
- Hunt EA, Hohenhaus SM, Luo X, *et al.* Simulation of pediatric trauma stabilization in 35 North Carolina emergency departments: identification of targets for performance improvement. *Pediatrics* 2006;117:641.
- Schmutz J, Manser T, Keil J, *et al.* Structured performance assessment in three pediatric emergency scenarios: a validation study. *J Pediatr* 2015;166:1498.
- Donoghue A, Nishisaki A, Sutton R, *et al.* Reliability and validity of a scoring instrument for clinical performance during Pediatric Advanced Life Support simulation scenarios. *Resuscitation* 2010;81:331–6.
- Tofil NM, Peterson DT, Wheeler JT, *et al.* Repeated versus varied case selection in pediatric resident simulation. *J Grad Med Educ* 2014;6:275.
- O'Leary F, McGarvey K, Christoff A, *et al.* Identifying incidents of suboptimal care during paediatric emergencies—an observational study utilising in situ and simulation centre scenarios. *Resuscitation* 2014;85:431–6.
- Katznelson JH, Mills WA, Forsythe CS, *et al.* Project CAPE: a high-fidelity, in situ simulation program to increase critical access hospital emergency department provider comfort with seriously ill pediatric patients. *Pediatr Emerg Care* 2014;30:397–402.

- 23 Roland D, Matheson D, Coats T, *et al.* A qualitative study of self-evaluation of junior doctor performance: is perceived 'safeness' a more useful metric than confidence and competence?. *BMJ Open* 2015;5:e008521.
- 24 Colthart I, Bagnall G, Evans A, *et al.* The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Med Teach* 2008;30:124–45.
- 25 Barnsley L, Lyon PM, Ralston SJ, *et al.* Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence. *Med Educ* 2004;38:358–67.
- 26 Regehr GL, Freeman RI, Robb AN, *et al.* OSCE performance evaluations made by standardized patients: comparing checklist and global rating scores. *Acad Med* 1999;74:S135–7.
- 27 Ilgen JS, Ma IW, Hatala R, *et al.* A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015;49:161–73.
- 28 Dellinger RP, Levy MM, Rhodes A, *et al.* Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013;41:580–637.